

INFORMATION DISCLOSURE CITATION PTO-1449		CUSTOMER NUMBER 44989	ATTORNEY'S DKT No. 0026-0072	APPLICATION No. 10/808,326	
			APPLICANT(S) Shioupyn SHEN		
			FILING DATE March 25, 2004	GROUP 2171	

U.S. PATENT DOCUMENTS						
EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE

FOREIGN PATENT DOCUMENTS							
EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation	
						Yes	No

OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)	
/GS/	Broder, Andrei et al., "Clustering the Web", http://research.compaq.com/src/articles/199707/cluster.html ; Systems Research Center, Palo Alto, California, February 4, 2004 (Print Date), 5 pages

EXAMINER /Garrett Smith/	DATE CONSIDERED 08/13/2007
---	---

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).

INFORMATION DISCLOSURE CITATION PTO-1449		CUSTOMER NUMBER 44989		ATTORNEY'S DKT No. 0026-0072	APPLICATION No. 10/808,326	
				APPLICANT(S) Shioupyn Shen		
				FILING DATE March 25, 2004	GROUP 2171	
U.S. PATENT DOCUMENTS						
EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE
/GS/	6,658,423 B1	12/2/2003	Pugh et al.	707	102	
FOREIGN PATENT DOCUMENTS						
EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation Yes No
OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)						
/GS/	Arvind Arasu et al.; Extracting Structured Data from Web Pages; Proceedings of the ACM SIGMOD 2003, 2003; 30 pages.					
/GS/	Brenda S. Baker; A Theory of Parameterized Pattern Matching: Algorithms and Applications (Extended Abstract); 25 th ACM STOC 1993; 1993; pages 71-80					
/GS/	Brenda S. Baker; On Finding Duplication and Near-Duplication in Large Software Systems; Proceedings of the 2 nd Working Conference on Reverse Engineering; 1995; 10 pages.					
/GS/	Andrei Z. Broder; On the Resemblance and Containment of Documents; Proceedings of the Compression and Complexity of Sequences; 1997; pages 1-9.					
/GS/	Krishna Bharat et al.; Mirror, Mirror on the Web: A Study of Host Pairs of Replicated Content; Proceedings of the 8 th International Conference on World Wide Web (WWW 1999); 17 pages.					
/GS/	Krishna Bharat et al.; A Comparison of Techniques to Find Mirrored Hosts on the WWW; Journal of the American Society for Information Science; 2000; 11 pages.					
/GS/	Krishna Bharat; The Connectivity Server: Fast Access to Linkage Information on the Web; Proceedings of the 7 th International Conference on the World Wide Web; 1998; 13 pages.					
/GS/	Andrei Z. Broder et al.; Min-Wise Independent Permutations; Proceedings of STOC; 1998; pages 630-659.					
/GS/	Sergey Brin et al.; Copy Detection Mechanisms for Digital Documents; Proceedings of the ACM SIGMOD Annual Conference, 1995; 12 pages.					
/GS/	Andrei Z. Broder et al.; Syntactic Clustering of the Web; Proceedings of WWW6, 1997; 13 pages.					
/GS/	James W. Cooper et al.; Detecting Similar Documents Using Salient Terms; Proceedings of the CIKM 2002; November, 2002; pages 245-251.					
/GS/	Edith Cohen et al.; Finding Interesting Associations without Support Pruning; Proceedings of the 16 th ICDE; 2000; 12 pages.					
/GS/	Abdur Chowdhury et al.; Collection Statistics for Fast Duplicate Document Detection; ACM Transactions on Information Systems; Vol. 20, No. 2; April 2002; pages 171-191.					
/GS/	Zhiyuan Chen et al.; Selectively Estimation for Boolean Queries; Proceedings of PODS 2000; 2000; 10 pages.					
/GS/	Jack G. Conrad et al.; Constructing a Text Corpus for Inexact Duplicate Detection; SIGIR 2004; July, 2004; pages 582-583.					
/GS/	Scott Deerwester et al.; Indexing by Latent Semantic Analysis; Journal of the American Society for Information Science; 1990; 34 pages.					
EXAMINER			DATE CONSIDERED			
/Garrett Smith/			08/13/2007			

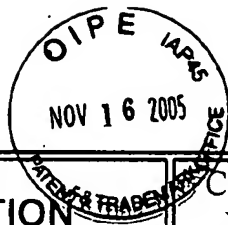
EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).

INFORMATION DISCLOSURE CITATION PTO-1449	CUSTOMER NUMBER 44989		ATTORNEY'S DKT No. 0026-0072		APPLICATION No. 10/808,326	
			APPLICANT(S) Shioupyn Shen			
			FILING DATE March 25, 2004		GROUP 2171	
U.S. PATENT DOCUMENTS						
EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE
FOREIGN PATENT DOCUMENTS						
EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation Yes No
OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)						
/GS/	Jeffrey Dean et al.; Finding Related Pages in the World Wide Web; Proceedings of the Eighth International World Wide Web Conference; 1999; pages 1-15.					
/GS/	Aristides Gionis et al.; Efficient and Tunable Similar Set Retrieval; Proceedings of SIGMOD 2001; 2001; 12 pages.					
/GS/	Taher H. Haveliwala et al.; Scalable Techniques for Clustering the Web; Proceedings of the 3 rd International Workshop on the Web and Databases (WebDB 2000); 2000; 6 pages.					
/GS/	Taher H. Haveliwala et al.; Evaluating Strategies for Similarity Search on the Web; Proceedings of the 11 th International World Wide Web Conference; May, 2002; 11 pages.					
/GS/	Khaled M. Hammouda et al.; Efficient Phrase-Based Document Indexing for Web Document Clustering; IEEE Transactions on Knowledge and Data Engineering; Vol. 16, No. 10, October 2004; pages 1279-1296.					
/GS/	Timothy C. Hoad et al.; Methods for Identifying Versioned and Plagiarised Documents; Journal of the American Society for Information Science and Technology; 2003; pages 1-18.					
/GS/	Sachindra Joshi et al.; A Bag of Paths Model for Measuring Structural Similarity in Web Documents; Proceedings of the 9 th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD 2003); August, 2003; pages 577-582.					
/GS/	Jon M. Kleinberg; Authoritative Sources in a Hyperlinked Environment; Journal of the ACM; 1999; 34 pages.					
/GS/	Aleksander Kotcz et al.; Improved Robustness of Signature-Based Near-Replica Detection via Lexicon Randomization; SIGKDD 2004; August, 2004; 6 pages.					
/GS/	Ravi Kumar et al.; Trawling the Web for Emerging Cyber-Communities; Computer Networks: The International Journal of Computer and Telecommunications Networks; 1999; 21 pages.					
/GS/	Udi Manber; Finding Similar Files in a Large File System; Proceedings of the 1994 USENIX Conference; 1994; 11 pages.					
/GS/	Athicha Muthitacharoen et al.; A Low-Bandwidth Network File System; Proceedings of the 18 th ACM Symposium on Operating System Principles (SOSP 2001); 2001; 14 pages.					
/GS/	Sean Quinlan et al.; Venti: A new Approach to Archival Storage; First USENIX Conference on File and Storage Technologies; 2002; 13 pages.					
EXAMINER			/Garrett Smith/		DATE CONSIDERED 08/13/2007	

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).

INFORMATION DISCLOSURE CITATION PTO-1449	CUSTOMER NUMBER 44989		ATTORNEY'S DKT No. 0026-0072		APPLICATION No. 10/808,326	
			APPLICANT(S) Shioupyn Shen			
			FILING DATE March 25, 2004		GROUP 2171	
U.S. PATENT DOCUMENTS						
EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE
FOREIGN PATENT DOCUMENTS						
EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation Yes No
OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)						
/GS/	Narayanan Shivakumar et al.; SCAM: A Copy Detection Mechanism for Digital Documents; Proceedings of the 2 nd International Conference on Theory and Practice of Digital Libraries; 1995; 9 pages.					
/GS/	Saul Schleimer; Winnowing: Local Algorithms for Document Fingerprinting; Proceedings of SIGMOD 2003; June 9-12, 2003; 10 pages.					
/GS/	Moses S. Charikar; Similarity Estimation Techniques from Rounding Algorithms; STOC 2002; May 19-21, 2002; 9 pages.					
/GS/	Gautam Pant et al.; Crawling the Web; 2003; 25 pages.					
EXAMINER			DATE CONSIDERED			
/Garrett Smith/			08/13/2007			

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).



SHEET 1 OF 1

INFORMATION DISCLOSURE CITATION PTO-1449	CUSTOMER NUMBER 44989	ATTORNEY'S DKT No. 0026-0072	APPLICATION No. 10/808,326				
		APPLICANT(S) Shioupyn Shen					
		FILING DATE March 25, 2004	GROUP 2171				
U.S. PATENT DOCUMENTS							
EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE	
FOREIGN PATENT DOCUMENTS							
EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation	
						Yes	No
OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)							
/GS/	Michael O. Rabin; Fingerprinting by Random Polynomials; Technical Report TR-15-81; Harvard University, 1981; 14 pages.						
EXAMINER				DATE CONSIDERED			
/Garrett Smith/				08/13/2007			

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).

INFORMATION DISCLOSURE CITATION PTO-1449		CUSTOMER NUMBER 44989		ATTORNEY'S DKT NO. 0026-0072		APPLICATION NO. 10/808,326	
				APPLICANT(S) Shioupyn Shen			
				FILING DATE March 25, 2004		GROUP 2165	
U.S. PATENT DOCUMENTS							
EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE	
/GS/	US 2005/0210043 A1	09/2005	Manasse				
/GS/	5,745,900	04/1998	Burrows				
/GS/	6,286,006 B1	09/2001	Bharat et al.				
/GS/	6,351,755	02/2002	Najork et al.				
/GS/	US 2003/0208761 A1	11/2003	Wasserman et al.				
/GS/	US 2006/0036684 A1	02/2006	Schwerk				
FOREIGN PATENT DOCUMENTS							
EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation	
						Yes	No
OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)							
/GS/	Office Action from U.S. Patent Application No. 11/094,791, dated July 23, 2007; 25 pages.						
/GS/	Arvind Jain et al.; U.S. Serial No. 11/094,791, filed March 31, 2005; "NEAR DUPLICATE DOCUMENT DETECTION FOR WEB CRAWLING"; 46 pages.						
EXAMINER				DATE CONSIDERED			
/Garrett Smith/				08/13/2007			

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).